Statistics and Bioinformatics Considerations

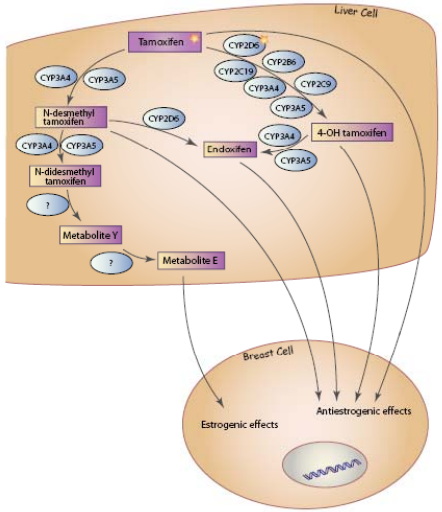
Lang Li, Ph.D.
COBRA
Associate Professor
Indiana University School of Medicine

Outlines

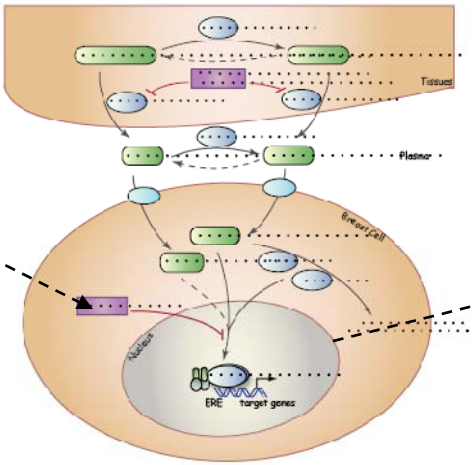
- Candidate Gene Association Studies
- Genome Wide Association Studies

Candidate Genes Association Study - Tamoxifen Data

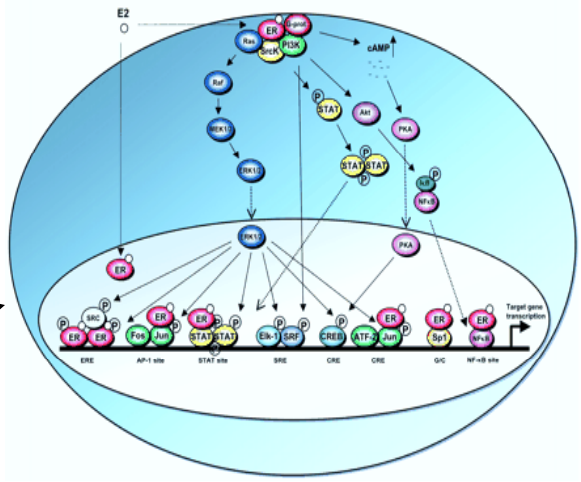
Tamoxifen Metabolism Pathway



Anti-Estrogen Pathway



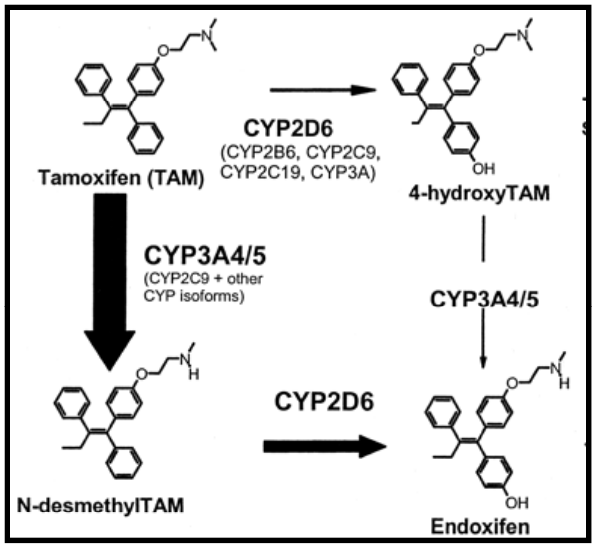
Estrogen Signaling and Regulation Pathway



Phenotypes: tamoxifen metabolites

Progression Free Survival
Hot Flashes, Bone Densities, Lipids

CYP2D6 Genetic Effect on Tamoxifen Metabolites in Patients with Breast Cancer



CYP2D6 allele	CYP2D6 Protein Function
3, 4, 5, 9	Knock-out
10, 17, 41	Intermediate
1, 2, 29, 35	Fully function
1xn, 2xn, 41xn	Ultra-Rapid

*3/*41
*17/*41
*4/*4
*41/*41
*4/*41
*10/*4
*10/*4xn
*35/*41
*1/*10
*10/*2
*35/*5
*10/*41
*2/*4
*1/*3
*2/*41xn
*2/*35
*1/*4
*5/*9
*1/*41
*1/*29
*1/*35
*35/*4
*1/*5
*2/*41
*41/*9
*1/*2
*2/*2
*1/*1
*2/*9
*10/*35
*1/*1xn
*2xn/*4
*1xn/*2
*41/*41xn
*1/*2xn

bi-allelic genotypes

- Hypothesis: the effect of CYP2D6 genotype on plasma NDM/endoxifen concentration
  - Phenotype: log(NDM / Endoxifen) ratio
  - Genotypes: 35 bi-allelic genotypes

Statistical Challenge

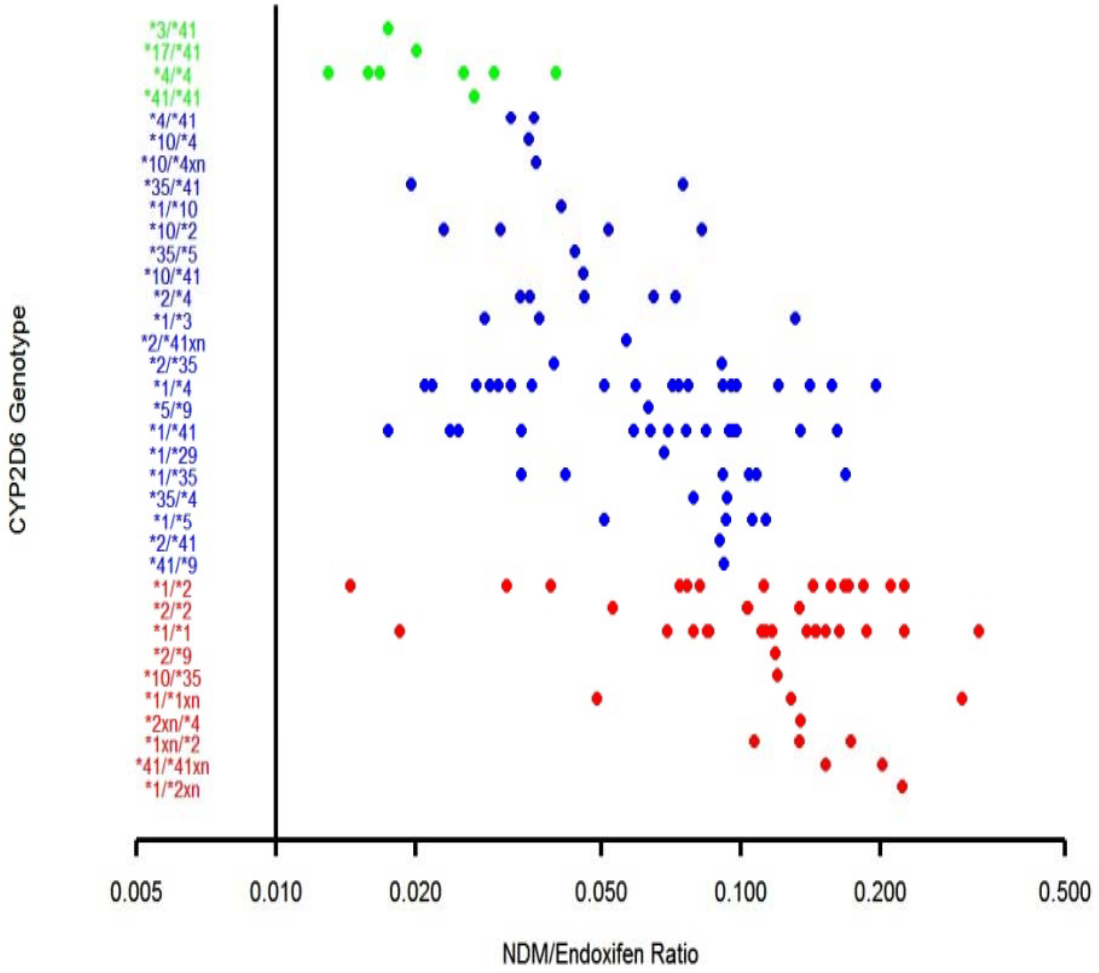
- **How do we test for possible associations between 35 CYP2D6 genotypes and a phenotype?**

  This results in 35*34/2 = 595 pair-wise comparisons e.g. *1/*1 vs *1/*4……

Possible Solutions from PGRN

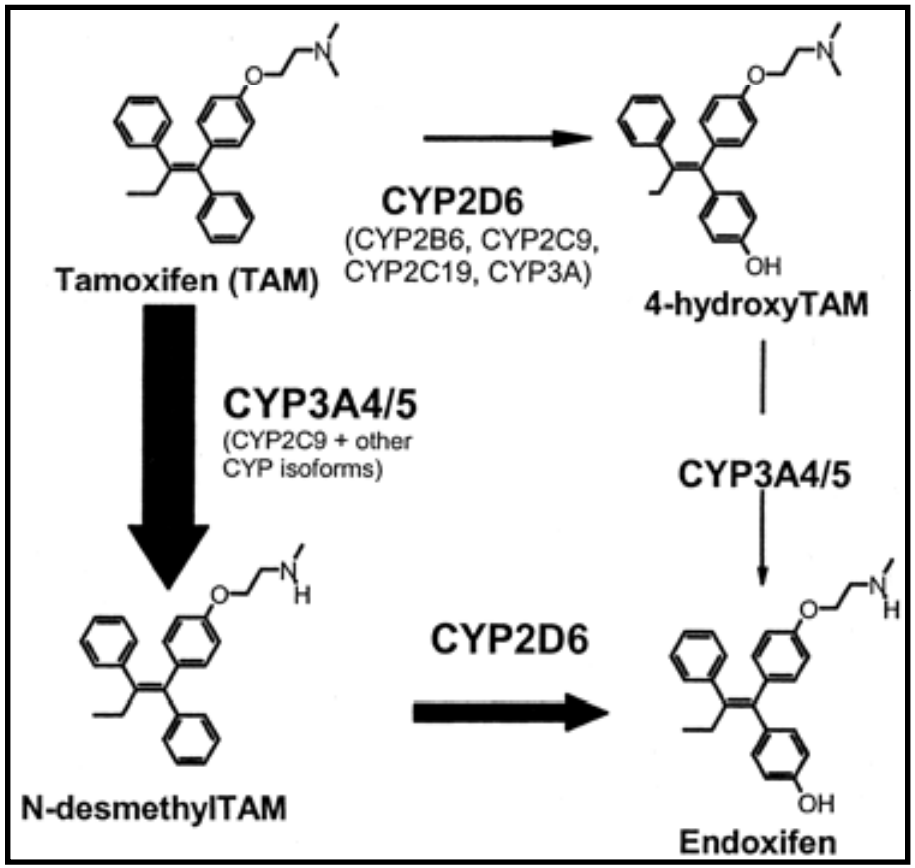
Multi-dimensional Reduction (MDR)	Marylyn D. Ritchie (Vanderbilt Univ., **PAT**)
Restricted Partition Method (RPM)	Robert Culverhouse (Washington Univ., **CREATE**)
Haplotype Score Tests	Daniel J. Schaid (Mayo Clinic, **PPII**)
Mixture Model	Lang Li (Indiana Univ., **COBRA**)

Using Mixture Model to Determine the CYP2D6 Genotype Clusters Based on the Phenotype



Clusters	P-value
1 vs 2	0.0008
2 vs 3	0.032
3 vs 4	0.143

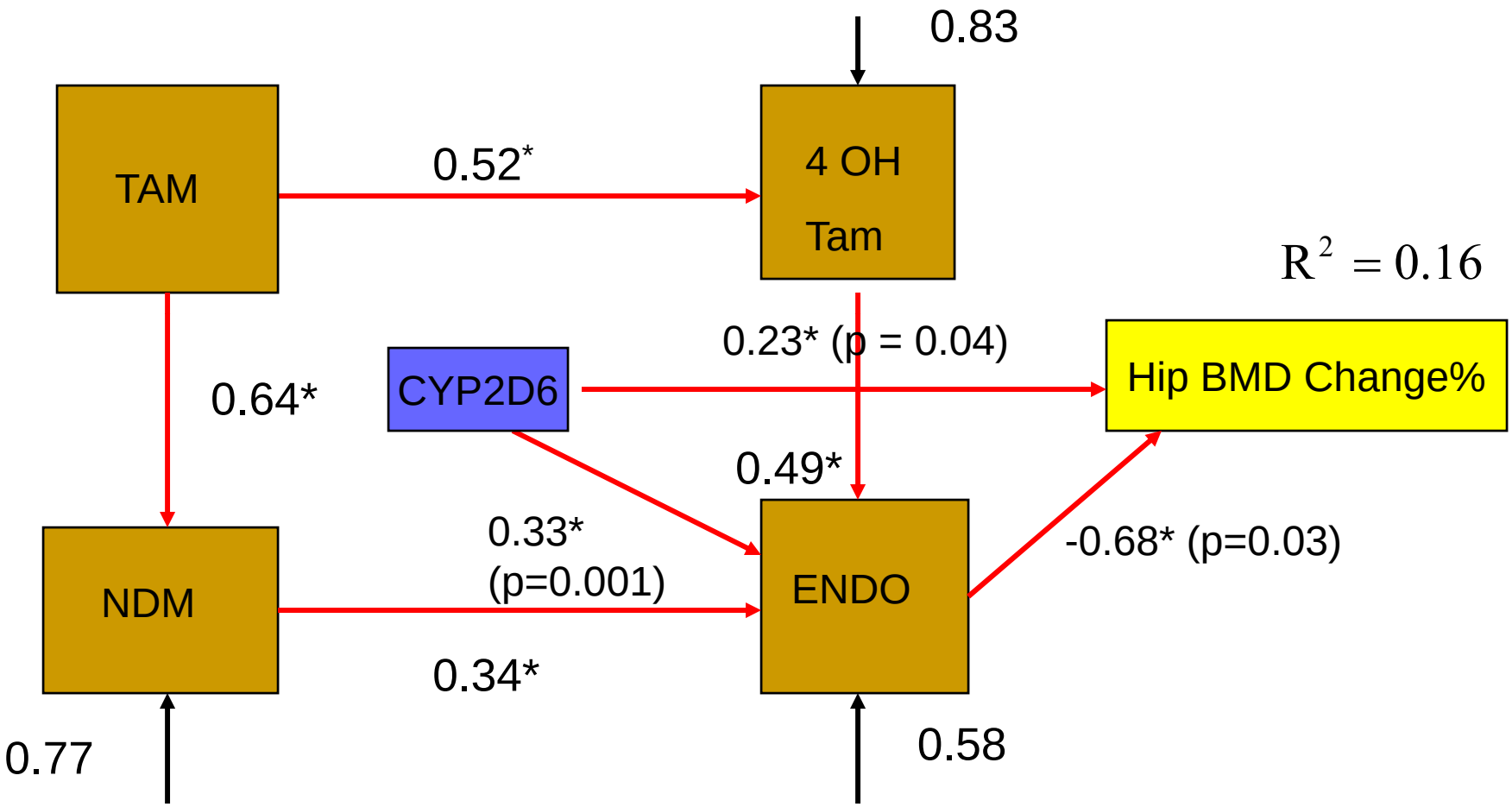
Genotype Can Be Used To Elucidate Mechanism: A Structure Equation Approach



BMD Change%

What is the contribution of CYP2D6 to different tamoxifen metabolites, and what are the pharmacodynamic consequences?

Using Structured Equation Modeling to elucidate genetic mechanisms: Tamoxifen effects on bone in pre-menopausal women



Genome Wide Association Studies

- Statistics
- Bioinformatics
- Systems Biology

GWAS Design: False Positive Control

- Liberal
  - Good for scientific discovery
  - False Discovery Rate
  - Per-comparison type I error
- Conservative
  - Good for clinical applications.
  - Family-wise type I error control.

GWAS Design: Replication

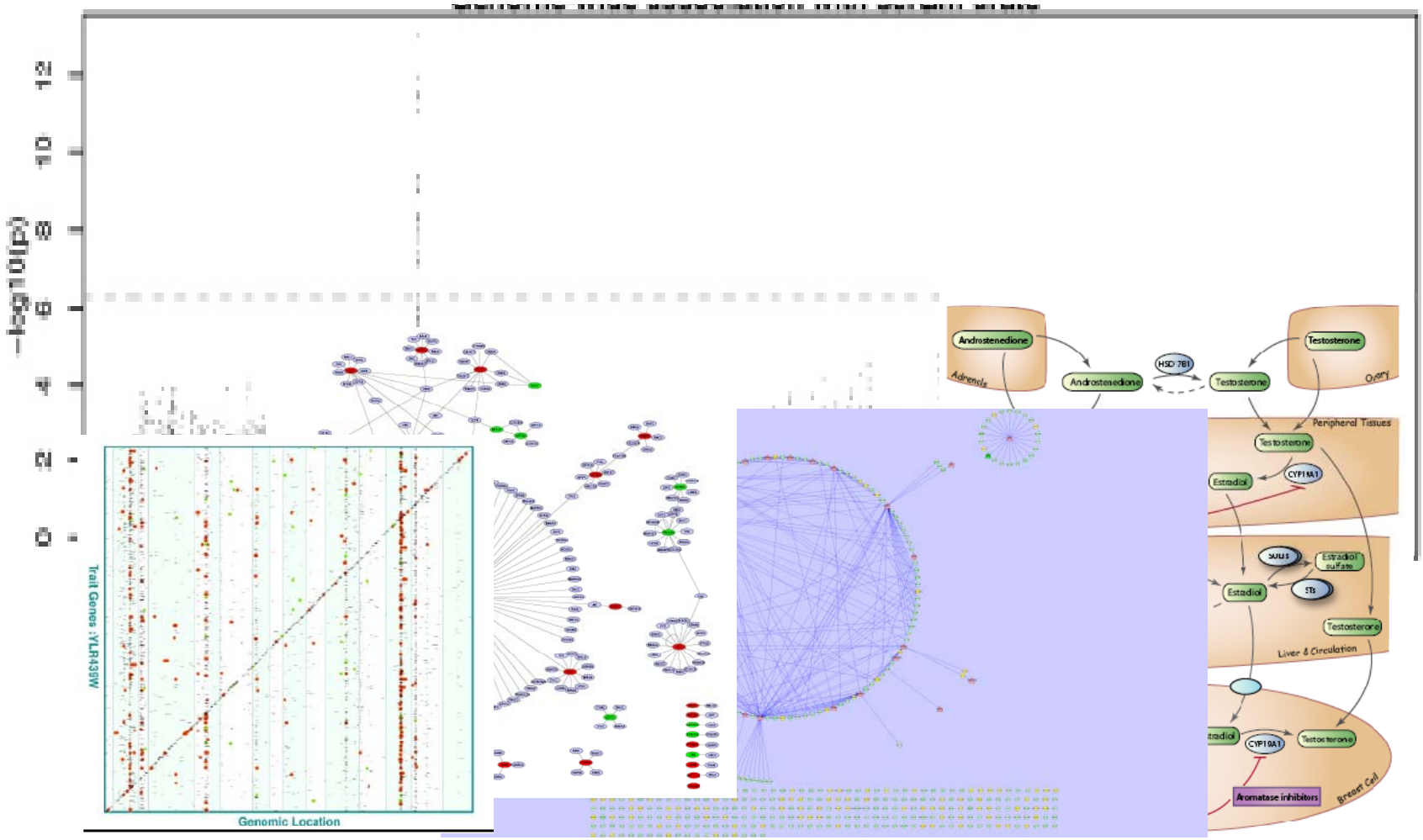
- Replications should preferably be conducted in independent data sets.
- Avoid the tendency to split one well-powered study into two less conclusive ones.
- The study designs of initial and replicated studies may differ in false positive control, the number of tested SNPs, and consequently the sample size.

(NCI-NHGRI Working Group, Nature, 2007)

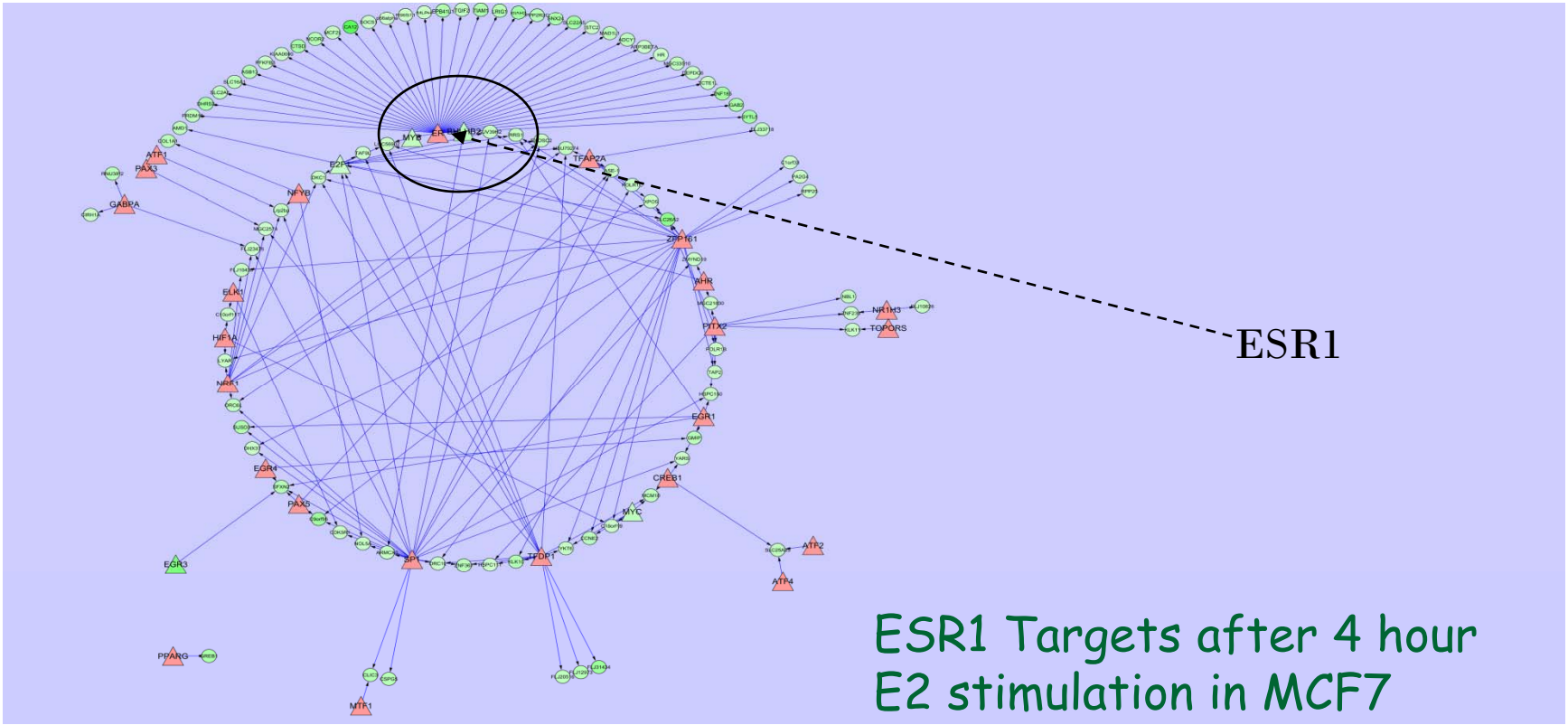
A Systems Biology Approach in GWAS

- **Why do we need systems biology?**
- **Critical to the Value and Interpretation of GWAS**
  - Current public domain signaling and pharmacology pathways are static and non-specific.
  - Most of these pathways are built upon the known mechanisms, which were obtained from small scale studies.
- **Every piece of high through-put data reflects important aspects of molecular networks.**

Data Output from a GWAS

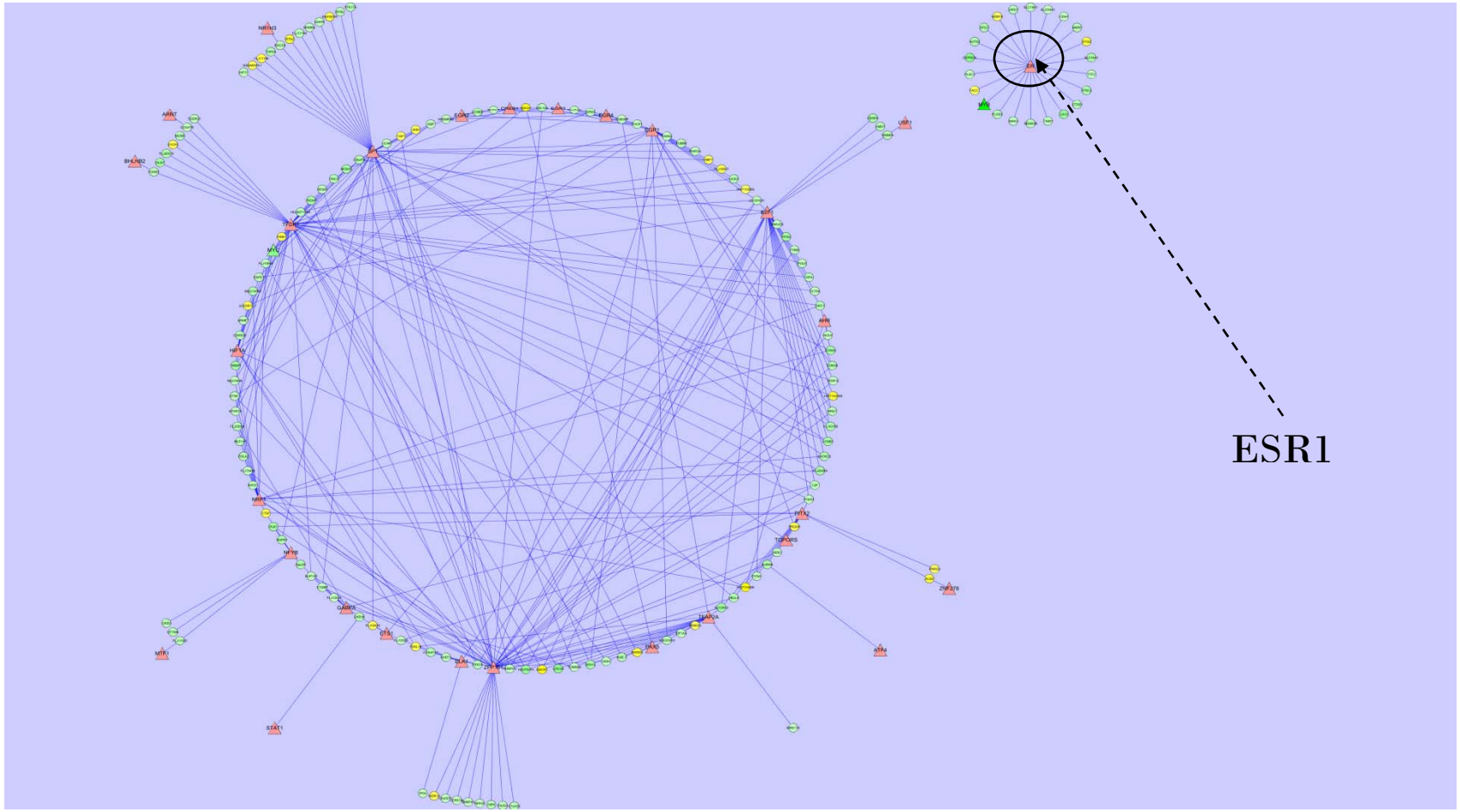


An Estrogen Regulation Network Model – An Integration of Gene Expression and ESR1 Tiling Array



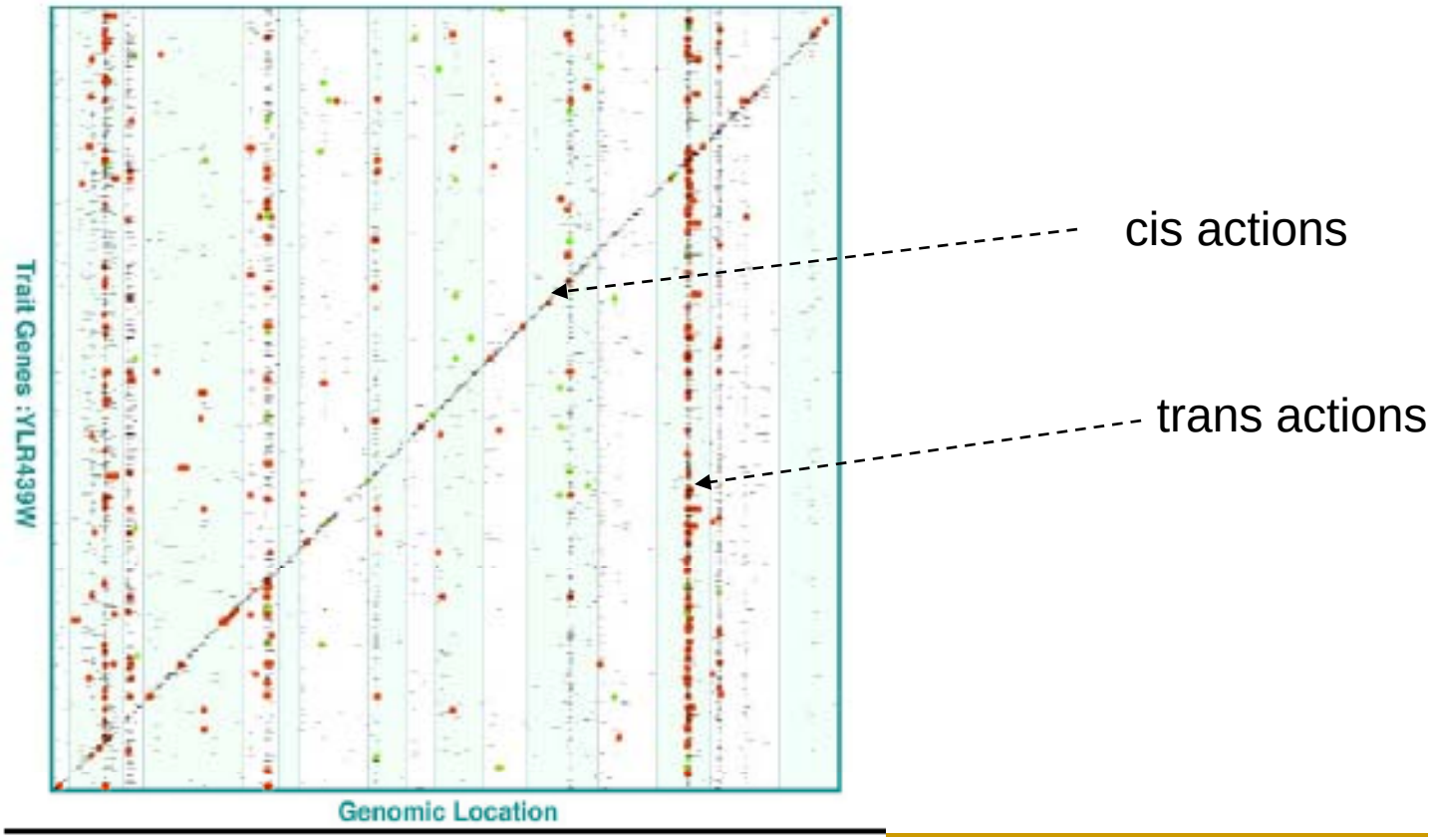
(Li et al. 2006, Shen et al. 2008)

ERα Targets After 24 hour E2 Stimulation

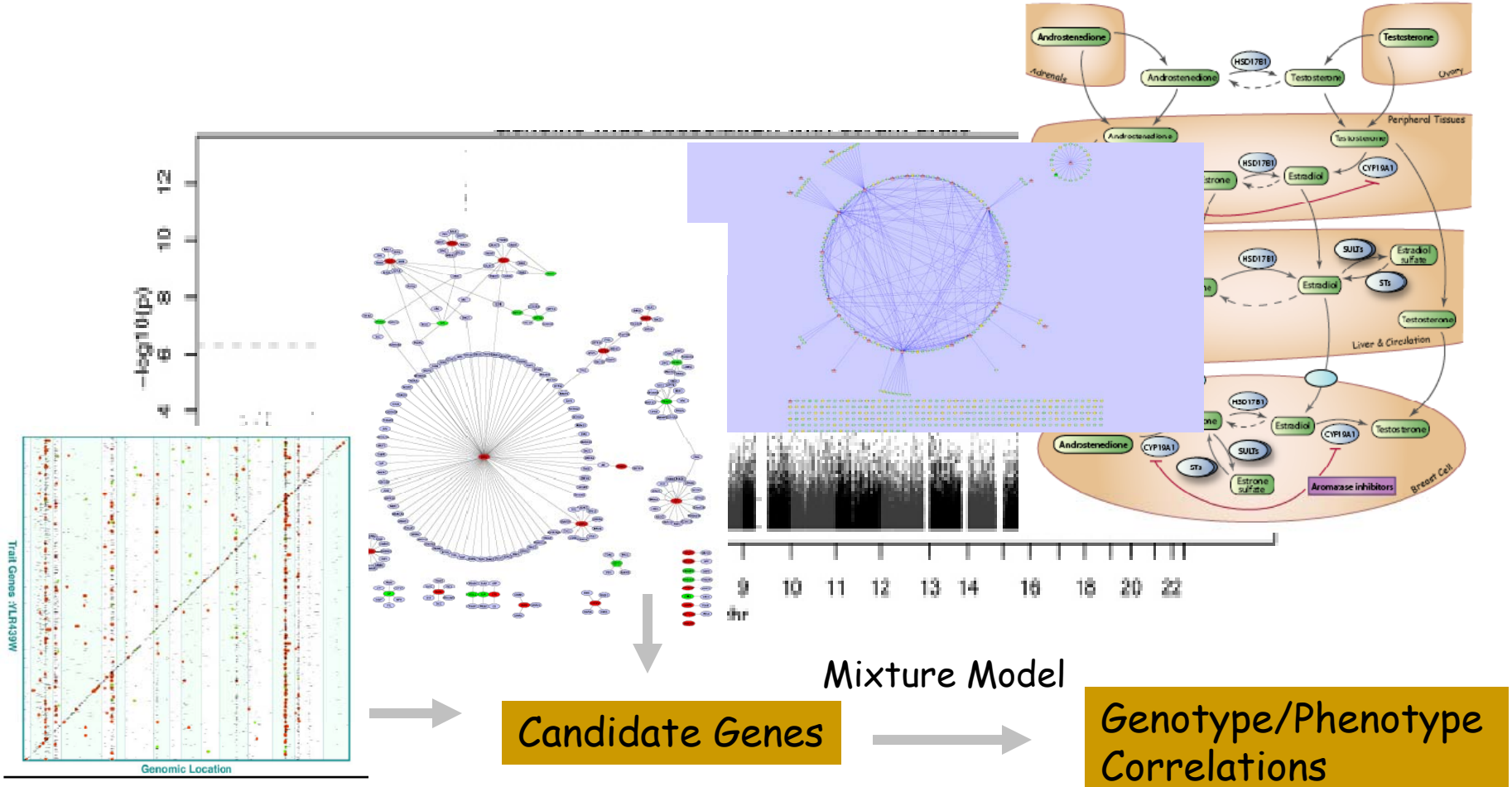


eQTL

SNP array/Gene Expression Interaction Network among breast cancer cell lines (eQTL)



An Integrated Approach in GWAS



Thank you!

GWAS – Statistics and Bioinformatics Tools

Statistics

- Plink
  http://pngu.mgh.harvard.edu/~purcell/plink/

- Merlin
  http://www.sph.umich.edu/csg/abecasis/

Bioinformatics

- SNP Annotation
  Plink
  http://pngu.mgh.harvard.edu/~purcell/plink/

- Gene-set enrichment Analysis
  - Ingenuity Pathway Analysis
    http://www.ingenuity.com/

  - Gene Set Enrichment Analysis
    http://www.broad.mit.edu/gsea/